Economics Working Papers (2002–2016) | Economics

7-18-2011

Hypothesis testing in linear regression when k/n is large

Gray Calhoun
Iowa State University, gcalhoun@iastate.edu

Hypothesis testing in linear regression when k/n is large

Abstract
This paper derives the asymptotic distribution of the F-test for the significance of linear regression coefficients as both the number of regressors, k, and the number of observations, n, increase together so that their ratio remains positive in the limit. The conventional critical values for this test statistic are too small, and the standard version of the F-test is invalid under this asymptotic theory. This paper provides a correction to the F statistic that gives correctly-sized tests under both this paper's limit theory and also under conventional asymptotic theory that keeps k finite. This paper also presents simulations that indicate the new statistic can perform better in small samples than the conventional test. The statistic is then used to reexamine Olivei and Tenreyro's results from "The Timing of Monetary Policy Shocks" (2007, AER) and Sala-i-Martin's results from "I Just Ran Two Million Regressions" (1997, AER).

Keywords
dimension asymptotics, F-test, ordinary least squares

Disciplines
Economics

Hypothesis Testing in Linear Regression when K/N is Large

Gray Calhoun

Working Paper No. 10041
December 2010

IOWA STATE UNIVERSITY
Department of Economics
Ames, Iowa, 50011-1070

Hypothesis Testing in Linear Regression when k/n is Large

Gray Calhoun*

Iowa State University, Department of Economics

July 18, 2011

Abstract

This paper derives the asymptotic distribution of the F-test for the significance of linear regression coefficients as both the number of regressors, k, and the number of observations, n, increase together so that their ratio remains positive in the limit. The conventional critical values for this test statistic are too small, and the standard version of the F-test is invalid under this asymptotic theory. This paper provides a correction to the F statistic that gives correctly-sized tests both under this paper's limit theory and also under conventional asymptotic theory that keeps k finite. This paper also presents simulations that indicate the new statistic can perform better in small samples than the conventional test. The statistic is then used to reexamine Olivei and Tenreyro's results from "The Timing of Monetary Policy Shocks" (2007, AER) and Sala-i-Martin's results from "I Just Ran Two Million Regressions" (1997, AER).

Keywords: Dimension Asymptotics, F-Test, Ordinary Least Squares

JEL Classification Numbers: C12, C20

*Address: 467 Heady Hall, Iowa State University; Ames, IA 50011; telephone: (515) 294-6271, email: gcalhoun@iastate.edu. Brendan Beare, Scott Borger, Graham Elliott, Nora Gordon, Seth Pruitt, Valerie Ramey, Yixiao Sun, and Allan Timmermann all gave me valuable feedback and advice, as did the participants at seminars at FDIC, Federal Reserve Bank of Kansas City, Iowa State University, UC San Diego, and Wayne State University, and two anonymous referees. Thank you all.

1 Introduction

Consider the linear regression model

$$y_t = x_t'\beta + \varepsilon_t \quad t = 1, \ldots, n$$

with x_t and ε_t uncorrelated. Under standard assumptions, the OLS estimator, $\hat{\beta}$, is consistent and asymptotically normal as n increases to infinity. This asymptotic distribution is the basis for most of the empirical research in economics, but as Huber (1973) has shown, it is unreliable unless k/n is close to zero; k is the number of regressors in the model. Huber proves that the OLS coefficient estimator is consistent and asymptotically normal when k increases with n, but only if $k/n \to 0$. In practice, k/n will always be positive and is sometimes large, so it is unclear whether the classic tests that exploit asymptotic normality are themselves reliable. This paper derives the asymptotic distribution of the F-test for arbitrary linear hypotheses about these coefficients under a more general limit theory that allows k/n to remain uniformly positive. The conventional F-test is asymptotically invalid under this limit theory, but despite this theoretical tendency to over-reject, will usually have close to its nominal size in practice.[1] Moreover, this paper derives a modification of the F-test that is asymptotically valid and demonstrates that this new test performs better than the unmodified F-test in practice.

This paper is not the first to study the asymptotic distribution of estimators like $\hat{\beta}$ as both n and k increase. Previous research has looked at the behavior of M-estimators as k increases, of Analysis of Variance (ANOVA) as the number of groups increases, and of instrumental variables estimators as the number of instruments increases. This research has followed two distinct paths. The first looks for the fastest growth rate of k that is compatible with standard consistency and asymptotic normality results; $k = o(n)$ is necessary for these results to hold but is often insufficient. The second approach looks for alternative asymptotic distributions of the coefficient estimators keeping k/n positive.

These increasing-k asymptotics were first introduced in the context of M-estimation; Huber (1973) argues that assuming k is fixed is unrealistic in practice. After proving that $k = o(n)$ is necessary for the OLS estimator to be consistent and asymptotically normal, Huber argues that this condition is likely needed by any tractable asymptotic theory and proves normality of the M-estimator of the coefficients of the linear regression model under the stronger condition that $k^3/n \to 0$. This rate was improved by Yohai and Maronna (1979), Portnoy (1984), and Portnoy (1985) to $k \log k/n \to 0$ for consistency and $(k \log k)^{1.5}/n \to 0$

[1] The F-test only performs well when using its finite-sample critical values. Tests based on the chi squared limit of the F statistic do not perform well in practice and should be avoided where possible.

for asymptotic normality. Further research has extended these results to other estimating functions (Welsh, 1989), nonlinear models (He and Shao, 2000), and estimation of the distribution of the errors (Chen and Lockhart, 2001; Mammen, 1996; Portnoy, 1986).

In econometrics, interest has focused instead on the properties of IV estimators with a fixed number of coefficients but an increasing number of instruments, l. Bekker (1994), building on earlier results by Anderson (1976), Kunitomo (1980), and Morimune (1983), studies the asymptotic behavior of Two-Stage Least Squares (2SLS) and variations of Limited Information Maximum Likelihood (LIML) in models with normal errors as l/n converges to a positive constant. These authors find that LIML is both consistent and asymptotically normal but that 2SLS is not. These results are extended to non-Gaussian errors by Hansen et al. (2008), Chao et al. (2008), and others. Koenker and Machado (1999) prove the consistency and asymptotic normality of GMM estimators with $l^3/n \to 0$. Stock and Yogo (2005), Chao and Swanson (2005), and Andrews and Stock (2007), among others, combine the many-instruments and the weak instruments literatures and argue that the relationship between the concentration parameter and l is more important than that between the number of observations and l. Anderson et al. (2010) establish some optimality properties for LIML in this setting. Han and Phillips (2006) study the limiting distributions of nonlinear GMM estimators with many weak instruments, and their approach allows for the estimators to converge to non-normal distributions.

Previous work on the F-test under increasing-k asymptotics has focused largely on ANOVA. Boos and Brownie (1995) find that the usual F-test is asymptotically invalid unless the design matrix is perfectly balanced (requiring an equal number of observations for each group) and propose a new Gaussian approximation for the statistic that gives an asymptotically valid test. This result is extended to two-way fixed-effects and mixed models (Akritas and Arnold, 2000); to allow for heteroskedasticity (Akritas and Papadatos, 2004; Bathke, 2004; Wang and Akritas, 2006); and to allow for additional covariates (Orme and Yamagata, 2006, 2007). See, for example, Fujikoshi et al. (2010) for many asymptotic results related to this literature. Anatolyev (forthcoming) studies the asymptotic performance of the Likelihood Ratio, LM, and F-tests under these asymptotics, imposing a different condition on the regressor matrix that rules out the unbalanced ANOVA applications just mentioned. Anatolyev shows that these three statistics behave differently; the LM and LR tests require a correction, but the F-test does not. We focus on the F-test alone in this paper, and find, consistent with the ANOVA literature, that it too requires a correction when the regressor matrix does not satisfy Anatolyev's conditions.

This research suggests that the standard test should behave poorly in finite samples unless the number of predictors is quite small. However, the F-test is known to have extremely good performance as a comparison of means, even when the errors are not normal. Scheffé (1959), for example, presents analytic and computational evidence that supports using the F-test even with asymmetric and fat tailed errors. Moreover, the simulations presented in some of the ANOVA papers themselves support using the naive F statistic instead of their proposed replacements. Akritas and Papadatos (2004), for example, simulate a 5% test with lognormal errors and find that the conventional F-test has size 0.04, while their proposed statistics have size 0.74 and 0.60, a moderate over-rejection.

These corrections have other undesirable features. The approximations do not hold under conventional, fixed-k asymptotics, forcing applied researchers to choose between two incompatible asymptotic approximations. Since k/n is always positive in practice, it is logical to use increasing-k limit theory by default, but the simulation evidence indicates that it performs poorly. Moreover, existing results only apply under strong restrictions on the matrix of regressors—assuming either an ANOVA structure or other inhibitive conditions—and so are not relevant for applied economic research.

This paper instead proposes a simple correction to the usual F statistic that gives a valid test under either conventional fixed-k or increasing-k asymptotics. When k is fixed, the correction disappears in the limit and our proposed statistic is asymptotically equivalent to the F-test. When k/n remains positive, the correction does not vanish and improves the size of the test statistic. The simulations presented in this paper indicate that this new statistic performs better than the conventional F-test and also outperforms a Gaussian test that is similar to those proposed in the ANOVA literature.

Since this statistic nests both the standard and nonstandard asymptotics, careful study of the correction can explain the F-test's strong performance in simulations. The magnitude of the correction depends on the excess kurtosis of the regression errors, ε_t, and on a particular feature of the design matrix of regressors. When the excess kurtosis is zero, no correction is necessary and the F-test is valid. If the excess kurtosis is not zero, the magnitude of the correction depends on the diagonal elements of the projection matrices for the unrestricted and restricted models—the restricted model is the model estimated under the null hypothesis. In practice, it is likely that the correction will be quite small and the naive F-test will perform reassuringly well, even if it is invalid. When the F statistic returns a value near the critical value for a specific test size, though, the correction can affect whether the test rejects or fails to reject the null hypothesis.

Finally, the use of this statistic is demonstrated through two applications—one for time series macroeconomic data and one for cross-sectional data. The first reexamines Olivei and Tenreyro's (2007) study, "The Timing of Monetary Policy Shocks," and finds further support for their conclusion that the effect of monetary policy on output has seasonal variation.

The second reexamines Sala-i-Martin's (1997) cross-country economic growth analysis and finds supporting evidence that additional variables beyond primary school education, GDP per capita, and life expectancy are correlated with a country's economic growth. These variables were singled out by Levine and Renelt (1992) and Sala-i-Martin (1997) as widely supported determinants of economic growth. The first example uses 144 observations to test 51 restrictions; the setup is a VAR with four equations and there are 51 restrictions on each of these equations. The second example tests uses 88 observations and tests 64 restrictions.

To reiterate, this paper derives a new statistic that can replace the F statistic in tests and works well for regression models with many regressors. The paper also explains the original F-test's strong performance in simulations and illustrates where it is likely to do poorly in applications. Section 2 discusses the new test statistic and studies its asymptotic distributions under the null and alternative hypotheses. Section 3 presents Monte Carlo evidence in favor of the statistic. Section 4 presents the empirical exercises. Section 5 concludes. The proofs are presented in the appendix.

2 Asymptotic Theory and Main Results

This section derives the asymptotic distribution of the F-test of the null hypothesis $R\beta = r$ for the linear equation

$$y_t = x_t'\beta + \varepsilon_t$$

as $q \to \infty$, $n \to \infty$ and q/n remains uniformly positive; q is the number of restrictions imposed by the null hypothesis. This limiting distribution implies that the F-test is not valid, and we present a new statistic, $\hat{G}$, that should be used instead of the F statistic. Comparing $\hat{G}$ to the quantiles from the $F(q, n-k)$ distribution yields an asymptotically valid test. Section 2.1 discusses the paper's notation and assumptions, Section 2.2 presents asymptotic theory and the new test statistic, and 2.3 studies the differences between the uncorrected and corrected statistics in more detail. Since the number of estimated coefficients is assumed to vary with n, a triangular array structure underlies all of this paper's theory. Unless otherwise indicated, all limits are taken as $n \to \infty$.

2.1 Assumptions

Since the number of restrictions imposed by the null hypothesis and the total number of predictors both increase with n, some assumptions take an unfamiliar form. They are, however, analogous to the usual assumptions that ensure the validity of the F-test under classical (fixed-k) asymptotic theory. The observations are required to be independent, the errors are required to be uncorrelated with the regressors and be homoskedastic, and the matrix $\mathbf{X}'\mathbf{X}$ is required to be uniformly positive definite.

The first assumption defines the behavior of the regressors and errors.

Assumption 1. *Define the random array $\{x_{n,t}, \varepsilon_{n,t}; t = 1, \dots, n\}$ and*

$$\mathbf{X}_n = (x_{n,1}, \dots, x_{n,n})'$$
$$\boldsymbol{\varepsilon}_n = (\varepsilon_{n,1}, \dots, \varepsilon_{n,n})'.$$

The elements $x_{n,t}$ are random k_n-vectors of regressors with bounded second moments, and there exist positive and finite constants b and B such that the eigenvalues of $n^{-1} \operatorname{E} \mathbf{X}_n'\mathbf{X}_n$ are less than B and greater than b for all n. For each n, the elements of the series $\{(x_{n,t}, \varepsilon_{n,t}); t = 1, \dots, n\}$ are independent and there are constants $r > 4$ and $B' > 0$ such that $\operatorname{E}|\varepsilon_{n,t}|^r < B'$ for all t and n. Moreover, $\operatorname{E}(\varepsilon_{n,t} \mid \mathbf{X}_n) = 0$ and $\operatorname{E}(\varepsilon_{n,t}^2 \mid \mathbf{X}_n) = \sigma^2 > 0$ for all t and n.

Assumption 1 restricts the errors to be strictly exogenous and conditionally homoskedastic, ruling out time series applications that use lagged dependent variables as predictors. The other details of this assumption could be relaxed. It would be straightforward, for example, to allow the array $\{x_{n,t}, \varepsilon_{n,t}\}$ to satisfy a less restrictive weak dependence condition than full independence, but the requirement that $\operatorname{E}(\varepsilon_{n,t} \mid \mathbf{X}_n) = 0$ is crucial.

The next assumption defines the relationship between $(\varepsilon_{n,t}, x_{n,t})$ and the dependent variable, $y_{n,t}$. The operator $|\cdot|_2$ denotes the Euclidean norm of an arbitrary vector in $\mathbb{R}^p$.

Assumption 2. *The dependent and independent variables are related through the equation*

$$y_{n,t} = x_{n,t}'\beta_n + \varepsilon_{n,t} \qquad t = 1, \dots, n \tag{1}$$

with $|\beta_n|_2 = O(1)$ as $n \to \infty$.

The assumption that $|\beta_n|_2 = O(1)$ ensures that the model does not asymptotically crowd out the error. If $|\beta_n|_2 \to \infty$ instead, the variance of $y_{n,t}$ would also increase to infinity, and in the limit (1) would behave as though there were no error. Such an asymptotic theory would obviously be of little practical value, as there is a substantial error term in real applications.

Also define the following notation. The OLS coefficient estimators are denoted $\hat{\beta}_n$ and the residuals are $\hat{\varepsilon}_{n,t}$. The null hypothesis of interest is

$$H_0: \qquad R_n\beta_n = r_n. \tag{2}$$

The next assumption embeds this hypothesis in a sequence of asymptotically well-behaved hypotheses.

Assumption 3. *$\{R_n\}$ is a sequence of $q_n \times k_n$ matrices of deterministic restrictions, each with full rank, and $\{r_n\}$ is a sequence of $q_n \times 1$ deterministic vectors.*

To minimize confusion between the F statistic, the F distribution, and the F-test, we denote the conventional F statistic for the hypothesis (2) as $\hat{F}_n$,

$$\hat{F}_n \equiv \frac{\sum_{t=1}^{n}\left(\hat{\varepsilon}_{n,t,0}^2 - \hat{\varepsilon}_{n,t}^2\right)/q_n}{\sum_{t=1}^{n}\hat{\varepsilon}_{n,t}^2/\left(n-k_n\right)}$$

with $\hat{\varepsilon}_{n,t,0}$ denoting the residuals from the restricted model.

2.2 Distribution of $\hat{F}$ and Asymptotic Correction

The theory proceeds in two steps. We first find the asymptotic distribution of the F-statistic as both q_n and n increase together. This distribution gives an unsatisfactory test statistic, but motivates an asymptotically equivalent test that performs better. Lemma 2.1 shows that $\hat{F}_n$ is approximately normal under the null hypothesis.

Lemma 2.1. *Suppose that Assumptions 1, 2, and 3 hold, that $q_n \to \infty$ and $k_n \to \infty$ with $\lim k_n/n < 1$, and that the null hypothesis (2) holds. Then*

$$\frac{\sqrt{q_n}}{\eta_n}\left(\hat{F}_n - 1\right) \xrightarrow{d} N(0,1), \tag{3}$$

with

$$\begin{aligned}
\eta_n^2 &= 2\left(1+c_n\right) + q_n^{-1}\sum_{t=1}^{n}\kappa_{n,t}D_{n,t}, \\
D_{n,t} &= \left(P_{n,tt}^* + c_n P_{n,tt} - c_n\right)^2, \\
P_n^* &= \mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}R_n'\left(R_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}R_n'\right)^{-1}R_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n', \\
P_n &= \mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n', \\
c_n &= \left(\frac{n-k_n}{n-k_n-2}\right)^2\frac{q_n+n-k_n-2}{n-k_n-4} - 1
\end{aligned}$$

and

$$\kappa_{n,t} = \mathrm{E}(\varepsilon_{n,t}^4 \mid \mathbf{X}_n)/\sigma^4 - 3.$$

Under the null hypothesis,

$$\sqrt{q_n}(\hat{F}_n - 1) = \frac{q_n^{-1/2}\boldsymbol{\varepsilon}_n'\left(P_n^* + c_nP_n - c_nI\right)\boldsymbol{\varepsilon}_n}{(n-k_n)^{-1}\boldsymbol{\varepsilon}_n'\left(I-P_n\right)\boldsymbol{\varepsilon}_n} + o_p(1),$$

so Lemma 2.1 follows from the asymptotic normality of the numerator; the denominator converges in probability to σ^2. The numerator can be shown to be asymptotically normal by an existing central limit theorem for quadratic forms, derived by Hall (1984) and de Jong (1987); the convergence in probability of the denominator to σ^2 follows from the same theorem. Details of the proof are presented in the appendix.

The terms η_n^2 and c_n need some explanation. The first term, η_n^2, is the asymptotic variance of $\sqrt{q_n}\hat{F}_n$ and comes from standard formulas for the variance of a quadratic form, applied to the numerator of $\hat{F}_n$. The second term, c_n, is chosen so that $2(1+c_n)/q_n$ is exactly equal to the variance of an $F_{q_n,n-k_n}$ random variable—as we will discuss shortly, $\eta_n^2 = 2(1+c_n)$ almost surely with normal errors, and in that case the variance of random variable on the left side of (3) equals one. For large n, c_n behaves like the simpler term $q_n/(n-k_n)$, but they can be noticeably different in moderate sample sizes.[2]

Lemma 2.1 implies that the standard F-test is asymptotically valid only if

$$q_n^{-1} \sum_{t=1}^{n} \kappa_{n,t} D_{n,t} \to 0$$

in probability, otherwise the asymptotic distribution of $\sqrt{q_n}\hat{F}_n$ is not pivotal. For example, observe that if $\varepsilon_{n,t} \sim N(0,\sigma^2)$, then $\hat{F}_n \sim F(q_n, n-k_n)$, and additionally $\sqrt{q_n}(\hat{F}_n - 1)$ converges to an $N(0, 2(1+\lim c_n))$ distribution (since $\kappa_{n,t} = 0$). This convergence implies that the critical values from the $F(q_n, n-k_n)$ converge to those of the normal$(0, 2(1+c_n))$ distribution as both q_n and $n-k_n$ increase together. Whenever

$$\sqrt{q_n}(\hat{F}_n - 1) \xrightarrow{d} N(0, 2(1+c_n)),$$

then, critical values from either the Gaussian or the $F(q_n, n-k_n)$ distribution can be used to test, implying that $q_n^{-1}\sum_{t=1}^{n} \kappa_{n,t} D_{n,t} \to 0$ is necessary for the naive F-test to be valid.

This sum converges to zero in three cases. First, if the excess kurtosis of the errors is zero the summation is identically zero, as the example with Gaussian errors illustrates. Second, if the design matrix, $\mathbf{X}_n$, is balanced, so that $P_{n,ss} = P_{n,tt}$ and $P^*_{n,ss} = P^*_{n,tt}$ for all s and t, then all of the elements $D_{n,t}$ are equal. In that case, since both P^*_n and P_n are idempotent matrices,

$$P_{n,tt} = n^{-1}\operatorname{trace}(P_n) = k_n/n \qquad a.s.$$

and

$$P^*_{n,tt} = n^{-1}\operatorname{trace}(P^*_n) = q_n/n \qquad a.s.,$$

[2]In simulations that we have not included in the paper, tests using $c_n = q_n/(n-k_n)$ over-reject by slightly less than the uncorrected F-test, but more than the proposed statistic.

so each $D_{n,t} = o_p(1)$ and the sum converges to zero.[3] Finally, if $q_n/n \to 0$ then each of the elements $P^*_{n,tt}$ converges to zero in probability and $c_n \to 0$, so each element of $D_{n,t}$ again converges to zero in probability. If none of those conditions are met, the sum does not vanish.

This Gaussian approximation suffers from some limitations. First, and most importantly, tests based on this approximation often perform worse than the naive F-test. Section 3 presents simulations that illustrate this point, and previous research is consistent with this claim. Akritas and Papadatos (2004), for example, run simulations for the F-test in a similar ANOVA application, and the naive F-test has lower Type-I error than their Gaussian alternatives. This Gaussian approximation also forces researchers to choose between asymptotic approximations. If k_n is fixed, the approximation implied by Lemma 2.1 does not hold because $q_n \hat{F}_n$ is asymptotically chi-square. Ideally, Lemma 2.1's approximation should contain the fixed-k_n result as a special case.

In light of these concerns, we propose rescaling the F statistic and the comparing that new statistic to the $F(q_n, n - k_n)$ critical values. Observe that Lemma 2.1 implies

$$\frac{\sqrt{2(1+c_n)q_n}}{\eta_n}(\hat{F}_n - 1) \xrightarrow{d} N(0, 2(1 + \lim c_n))$$

under Assumptions 1–3. As discussed, the normal$(0, 2(1+c_n))$ distribution and the $F(q_n, n - k_n)$ distribution are related: any sequence of random variables G_n that satisfies $\sqrt{q_n}(G_n - 1) \xrightarrow{d} N(0, 2(1 + \lim c_n))$ is approximately $F(q_n, n - k_n)$ as well when both q_n and $n - k_n$ are large. Theorem 2.2 defines G_n as

$$G_n = \eta_n^{-1}\sqrt{2(1+c_n)}(\hat{F}_n - 1) + 1,$$

and exploits this relationship. This random variable could form the basis of an infeasible test statistic instead of $\hat{F}_n$. If G_n exceeds the $(1 - \alpha)$ quantile of the $F(q_n, n - k_n)$ distribution, the test would reject.

Theorem 2.2. *Suppose that the conditions of Lemma 2.1 hold but q_n and k_n may be bounded. Define the random variable*

$$G_n = v_n \hat{F}_n + (1 - v_n), \qquad v_n = \frac{\sqrt{2(1+c_n)}}{\eta_n}.$$

Under (2), $\mathbf{P}[G_n > z_{n,\alpha}] \to \alpha$; each $z_{n,\alpha}$ is the $(1 - \alpha)$ critical value for the F distribution with q_n and $n - k_n$ degrees of freedom.

[3] Anderson et al. (2010) discuss this condition in the context of LIML and give an example of asymptotic balance. Independent Standard Normal regressors, for example, would ensure asymptotic balance

This asymptotic approximation suffers from none of the drawbacks of the first. The simulations presented in Section 3 demonstrate that the feasible test statistic based on G_n performs at least as well as the F-test. Moreover, this asymptotic result holds whether or not q_n and k_n increase, so researchers do not have to choose between asymptotic theories. When q_n is bounded, $\eta_n^2 \to 2(1 + \lim c_n)$ in probability, so the correction vanishes.

The correction term v_n is a variance correction. When the innovations have positive excess kurtosis, the variance of the F statistic is larger than predicted by the $F(q_n, n - k_n)$ distribution. For small values of q_n, the variance is only slightly larger, but for large values of q_n relative to $n - k_n$, this discrepancy can affect test results. Applying the proposed correction, v_n, simply rescales the F statistic so that its true variance matches that of the F distribution.

This correction must be estimated to make testing feasible; in particular, η_n^2 is an unknown random variable. The next lemma gives an estimator for η_n^2. For simplicity, we assume that all of the errors have equal excess kurtosis. Our argument and proof hold for non-constant kurtosis as well, but the formula becomes more complicated. Even under equal kurtosis, the natural estimators of that kurtosis,

$$\frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_{n,t}^4 / \hat{\sigma}^2 - 3$$

or

$$\frac{1}{n - k_n} \sum_{t=1}^{n} \hat{\varepsilon}_{n,t}^4 / \hat{\sigma}^2 - 3,$$

are inconsistent, since $\hat{\beta}_n$ is inconsistent itself. However, by expanding $(\varepsilon_{n,t} - x_{n,t}'(\hat{\beta}_n - \beta_n))^4$, we can find and remove the asymptotic bias of the first estimator and show that the variance of this new estimator converges to zero. Lemma 2.3 gives the necessary correction and provides a consistent estimator for η_n^2.

Lemma 2.3. *Suppose that the conditions of Theorem 2.2 hold and that $\kappa_{n,t} = \kappa_n$ for all t, and define*

$$\hat{\kappa}_n = n^{-1} \textstyle\sum_{t=1}^{n} \dfrac{\hat{\varepsilon}_{n,t}^4 / \hat{\sigma}^4 - w_{n,1}}{w_{n,2}} - 3, \tag{4}$$

with

$$w_{n,1} = n^{-1} \sum_{t=1}^{n} \left(6P_{n,tt} - 15P_{n,tt}^2 + 12P_{n,tt}^3 - 3 \sum_{s=1}^{n} P_{n,st}^4 \right) \tag{5}$$

and

$$w_{n,2} = n^{-1} \sum_{t=1}^{n} \left(1 - 4P_{n,tt} + 6P_{n,tt}^2 - 4P_{n,tt}^3 + \sum_{s=1}^{n} P_{n,st}^4 \right). \tag{6}$$

Then

$$2(1 + c_n) + (\hat{\kappa}_n / q_n) \sum\nolimits_{t=1}^{n} D_{n,t} = \eta_n^2 + o_p(1). \tag{7}$$

If, in addition, the null hypothesis is true then (7) *also holds with the residuals from the restricted model replacing each* $\hat{\varepsilon}_{n,t}$ *and* $P_{n,ts} - P_{n,ts}^*$ *replacing each instance of* $P_{n,ts}$ *in* (5) *and* (6)*.*

Although the residuals from either the restricted or unrestricted models can be used to estimate η_n^2, we recommend that the restricted residuals be used in practice. Since there are fewer coefficients, the kurtosis can be estimated more precisely if the null is imposed.

The asymptotic distribution of the feasible test statistic is an immediate corollary.

Corollary 2.4. *Suppose that the conditions of Theorem 2.2 are satisfied, let* $\hat{v}_n$ *be a consistent estimator of* v_n*, and define*

$$\hat{G}_n = \hat{v}_n \hat{F}_n + (1 - \hat{v}_n). \tag{8}$$

If (2) *holds,* $\mathbf{P}[\hat{G}_n > z_{n,\alpha}] \to \alpha$*.*

In theory, this adjustment can improve either the size or the power of the F-test, depending on whether the excess kurtosis of the errors is positive or negative. However, the consequences of erroneously shrinking or expanding the test statistic are quite different. The first reduces the test's power, and the second increases the test's size. Since trying to exploit negative estimates of the excess kurtosis will cause the test to over-reject if those negative estimates are wrong, we suggest that researchers only adjust the test statistic if the estimate is positive. This practice amounts to using the statistic

$$\hat{G}_n = \min\{\hat{v}_n, 1\} \hat{F}_n + (1 - \min\{\hat{v}_n, 1\})$$

for testing.

The preceding discussion has focused on the validity of the F-test and our proposed alternative, but we also care about the power of these tests. Since the asymptotic theory for G_n and $\hat{G}_n$ is founded on asymptotic normality, it is relatively easy to derive their distributions under local alternatives of the form

$$R_n\beta_n = r_n + \delta_n. \tag{9}$$

Corollary 2.6 shows that the test based on G_n has nontrivial power if $\delta_n'\delta_n = O(q_n^{1/2}/n)$ and is consistent if δ_n converges to zero more slowly. An important special case is if $R_n = I_k$, $r_n = \mathbf{0}$, and $\delta_n = (1, 0, \dots, 0)'$—i.e. testing for the joint significance of the regression when one of coefficients is nonzero. In that case, the test has unit power asymptotically.

Lemma 2.5. *Suppose that the conditions of Lemma 2.1 hold but that the alternative hypothesis (9) holds with $\delta_n'\delta_n \sim q_n^{1/2}/n$. Then*

$$\frac{\sqrt{q_n}}{\eta_n}\left(\hat{F}_n - 1\right) - \theta_n \to N(0,1) \tag{10}$$

in distribution, with

$$\theta_n \equiv \frac{\delta_n'(R_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}R_n')^{-1}\delta_n}{\sigma^2\eta_n^{-1}\sqrt{q_n}} \sim 1.$$

The behavior of $\hat{F}_n$ under local alternatives is sufficient to describe the first-order local power of $\hat{G}_n$. Exploring the higher-order behavior of G_n or $\hat{G}_n$ is beyond the scope of this paper. Consistency of the test is an immediate corollary.

Corollary 2.6. *Suppose that the conditions of Corollary 2.4 hold, but that the alternative hypothesis* (9) *holds with $\delta_n'\delta_n \sim 1$. Then $\mathbf{P}[\hat{G}_n > z_{n,\alpha}] \to 1$.*

2.3 Behavior of the F-statistic

Lemma 2.1 and Theorem 2.2 show that the F-test is invalid and propose a corrected replacement test statistic, but if the correction, v, is near one the replacement may be unnecessary because the F-test may do well in practice.[4] This section looks at the relationship between the magnitude of the infeasible correction, v, and the size of the uncorrected F-test. If the effect of v on the F-test were small, researchers might prefer to use the uncorrected F-test out of convenience. However, this section suggests that the new test is preferable. In this section, we assume that the fourth moments of the errors are identical, and the correction simplifies considerably:

$$v = 2(1+c) + \kappa\bar{D}, \qquad \bar{D} = \sum\nolimits_{t=1}^{n} D_t/q. \tag{11}$$

[4] In the remainder of the paper, we drop the stochastic array notation since we are presenting finite-sample results.

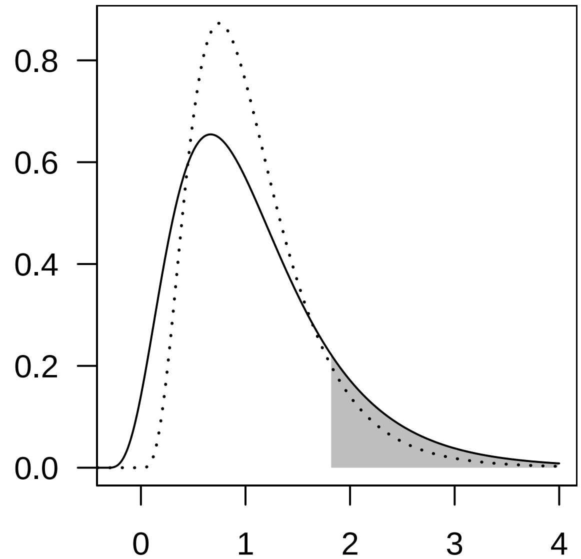

Figure 1: $F(15, 70)$ density (dotted line) and the density $f_{\hat{F}}(\cdot)$ given in Equation (2.3) for $n = 100$, $k = 30$, $q = 15$, and $v = 0.5$ (solid line). The shaded region is the area to the right of the 0.90 quantile of the $F(15, 70)$ distribution.

This paper's asymptotic theory implies that G has approximately an $F(q, n - k)$ distribution; in this subsection, we assume that G has that exact distribution, and derive the implied density of the uncorrected F-statistic. We then calculate the mass of that density above the 0.90, 0.95, and 0.99 quantiles of the F-distribution for different values of n, k, q, and $\kappa\bar{D}$. This mass gives an indication of the extent to which the naive F-test over-rejects; if 5% of the distribution of the F-statistic lies to the right of the 0.99 quantile, then the true size of a nominal 1% test is 5%. These calculations are approximate since G does not have this distribution in finite samples, but they do allow the impact of $\kappa\bar{D}$ to be isolated.

If the distribution of G is known and v is fixed, it is straightforward to derive the distribution of $\hat{F}$. Suppose that $f_{q,n-k}(\cdot)$ is the density of the $F(q, n - k)$ distribution. Since

$$\hat{F} = (G + v - 1)/v,$$

the density of $\hat{F}$, denoted $f_{\hat{F}}(\cdot)$, is trivially calculated to be

$$f_{\hat{F}}(x) = v \cdot f_{q,n-k}(vx + 1 - v).$$

Figure 1 plots a representative graph of the densities $f_{q,n-k}$ and f_F for $n = 100$, $k = 30$, $q = 15$, and $v = 0.5$. We see that both densities are centered at one, and that the density of $\hat{F}$ is more dispersed than that of G. The area under the density of $\hat{F}$ that lies to the right of the 90% critical value of the F-distribution is shaded. This area is equal to 0.15, so the

F-test would over-reject.

To better understand these size distortions, we calculate this area for different values of n, k, q, and $\kappa\bar{D}$. We let n be 100 or 500, k be $n/10$ or $n/2$, and q be 1, $k/2$, or $k-1$. We consider all values of $\kappa\bar{D}$ between –1 and 3 and consider tests of size 1%, 5%, and 10%. The size distortions are plotted in Figure 2. The plots are arranged in a grid, and each plot in the grid contains three curves that depict the area for a different quantile. The solid line depicts $n = 100$ and the dotted line $n = 500$. The true area is the vertical axis and the value of $\kappa\bar{D}$ is the horizontal axis. When $\kappa\bar{D}$ is zero, no correction is necessary and each area is equal to the corresponding area of the F density function. In each plot, the top curve presents the values for a test of size 10%, the middle curve 5%, and the bottom curve 1%.

The broad patterns are the same for each test size and do not depend heavily on n, although the size distortions are slightly larger for larger values of n. These distortions are, obviously, smaller when $\kappa\bar{D}$ is near zero and increase as $\kappa\bar{D}$ increases. For $\kappa\bar{D}$ near three, the true size of each test is roughly 5 percentage points higher than the test's nominal size. For any values of n and k, the distortions increase with q, and for any values of n and q, the distortions do not seem to vary with k.

3 Monte Carlo comparison

Although the asymptotic properties of the new statistic, $\hat{G}$, are superior to the conventional F-test, finite-sample properties are more important. One possible concern about this new statistic is the need to estimate the regression errors' kurtosis; if that estimate is poor, the test may perform worse than the uncorrected F-test. This section presents a Monte Carlo simulations studying the size of the test statistic based on $\hat{G}$ and shows that the new statistic performs better.[5]

To study the size of these tests, we simulate data for the equation

$$y_t = \beta_0 + \beta' x_t + \varepsilon_t$$

where β_0 and β both equal zero; each element of x_t is drawn from the same distribution, either Normal or Cauchy; and ε_t is drawn from Student's t distribution with 5 or 30 degrees of freedom, or from the Exponential distribution (centered to have mean zero). These distributions let us study the effects of heavy tails and asymmetric errors on the test statistics. Other simulations, not presented here, indicate that the imbalance term, $\sum D_t/q$, is high for Cauchy and small for Normal regressors.

[5] These simulations were conducted in R version 2.11.1 (R Development Core Team, 2010).

Each simulation constructs four tests of the null hypothesis

$$\beta_1 = \cdots = \beta_q = 0,$$

the usual F-test, the new test based on $\hat{G}$, the Wald test, and a test based on the normal approximation in Lemma 2.1. The variance, η^2, used by $\hat{G}$ and the normal approximation is estimated using the residuals from the restricted regression.[6] Moreover, if the variance estimate is less than the population variance of the $F_{q,n-k}$ random variable, we use the population variance instead (this amounts to restricting the correction term, $\hat{v}$, to be less than one), as discussed in Section 2.2. We let n be 100 or 500, k be $n/10$ or $n/2$, and q be 1, $k/2$, or $k-1$. We only study the size of nominal 5% tests; different sized tests give similar results and are not reported.

Tables 1 and 2 contain the results of these simulations. Each entry lists the percentage of the 5000 simulations that reject the null hypothesis. The naive F-test and test based on $\hat{G}$ both perform well for almost all of the simulations we consider; their simulated size is very close to the nominal size of 5%. The exception is for Cauchy predictors with t_5 and Exponential errors. With t_5 errors, the F-test over-rejects by roughly 5 percentage points unless a single restriction is tested, and the amount of over-rejection increases with n, k/n, and q. The simulated size of $\hat{G}$, on the other hand, remains much closer to its nominal size and over-rejects by only one or two percentage points. For exponential errors, both statistics perform slightly worse, but the $\hat{G}$-test over-rejects by at most two percentage points.

The other statistics perform worse. The test using the normal approximation over-rejects by several percentage points more than the naive F-test for almost all of the parameter values studied. The Wald test performs by far the worst. It performs acceptably when a single restriction is being tested, but when k/n or q is large its size deteriorates, even when the naive F test performs well. For the simulations that the naive F test does not preserve size, Cauchy regressors with q large, the Wald test can over-reject by over 15 percentage points.

Since our theoretical results depend crucially on assuming homoskedasticity, we also examine the tests' performance under heteroskedasticity. We generate data from the same equation and same parameters, but now scale ε_t by $1 + F_x^{-1}(x_{2,t})$, with $F_x^{-1}(\cdot)$ the inverse of the CDF of the stochastic regressors.[7] To save space, we only present results for Cauchy regressors. We also use the robust variance-covariance matrix to construct the Wald test,

[6] Unreported simulations indicate that the test performs worse if the unrestricted regression is used to estimate the variance.

[7] Scaling by $1 + x_{2,t}$, which might be more standard, introduces much too much heteroskedasticity for any of the statistics to be reliable when we use Cauchy regressors. For Normal regressors, we get qualitatively similar results.

but leave the other statistics unchanged.

Table 3 presents the simulation results. For $q = 1$, all of the statistics perform acceptably well, but the F-test and normal test perform badly for larger values of q, over-rejecting by about 5 to 10 percentage points. The F-test now performs slightly worse than the normal test (about one or two percentage points), which did not happen with homoskedasticity. Surprisingly, the robust Wald test performs well in these simulations, with a size of about 7% across the experiments. The $\hat{G}$-test again performs the best—systematically better than the robust Wald test by about a percentage point. This performance is also surprising, especially in light of the poor performance of the F-test, but is reassuring and indicates that the test should be reliable in the presence of mild undiagnosed heteroskedasticity. Obviously, these simulations only examine one form of heteroskedasticity, and researchers should proceed cautiously if they suspect there is strong heteroskedasticity.

Generally, these simulations demonstrate that the proposed statistic, $\hat{G}$, preserves size well in finite samples, even for models with many regressors and hypotheses with many restrictions, and even when the homoskedasticity assumption is violated. When q is large, the restricted model uses only a few regressors, $k - q$. In that case, the kurtosis can be estimated precisely, and $\hat{G}$ should be expected to perform well. On the other hand, when q is small, k_o is large but $\bar{D}$ is small. In this case, the kurtosis may be estimated poorly, but its estimate has only a small effect on the final statistic. In the few cases where the new statistic over-rejects, it does so less than the alternative statistics.

4 Empirical Exercise

This section presents two empirical studies that illustrate our new test statistic.[8] The first is a macroeconomic application based on Olivei and Tenreyro's (2007) study of monetary policy shocks and the second is a cross-sectional application based on Levine and Renelt's (1992), Sala-i-Martin's (1997), and Sala-i-Martin et al.'s (2004) studies of economic growth. Although this paper's theory has not yet been extended to time series applications with lagged dependent variables, which is the econometric model used by Olivei and Tenreyro, their study is a natural application of this paper's statistic and it is unlikely that the form of the test statistic based on $\hat{G}$ will need to change to be appropriate in such applications.

[8]Both empirical exercises were carried out using R version 2.11.1 (R Development Core Team, 2010)

4.1 Monetary Policy Shocks

Macroeconomic models often impose rigidities in price and wage contracts so that monetary policy has an effect on real economic variables, output and unemployment in particular. Taylor (1980) and Calvo (1983) pioneered models with many agents who set wages or prices rationally but infrequently; the length for which the price is set is exogenous, and the agents set these prices at different, staggered, times. In this framework, aggregate prices and wages do not change instantaneously, and these frictions can cause agents to change their consumption and labor supply in response to changes by the Federal Reserve to the money supply or interest rate.

Olivei and Tenreyro (2007) argue that these models could be missing important seasonal effects in the price rigidity that would cause monetary policy to have a stronger effect at some times of the year than others. Sticky price models are often motivated by citing union wage negotiations and other similar contracts. Olivei and Tenreyro cite survey evidence that most firms renegotiate these contracts in the fourth quarter of the calendar year, and that these changes are enacted in the first quarter of the next year. Consequently, actions by the Federal Reserve could have less impact in the first and fourth quarters, when wages and prices are the most flexible.

Olivei and Tenreyro formalize this argument in two ways. They develop a variation of a Calvo sticky-wage Dynamic Stochastic General Equilibrium (DSGE) model that allows the probability of wage renegotiation to vary over the year. They also estimate a structural vector autoregression (SVAR) using GDP, the GDP deflator, an index of commodity prices, and the Federal Funds rate, and allow the coefficients of this model to be different in each quarter.[9] They find that the impulse response functions of the DSGE model match those estimated from the SVAR and show a more pronounced effect from monetary policy shocks in the second and third quarters.

This section will focus on one aspect of Olivei and Tenreyro's work: whether the VAR coefficients are truly different across quarters. Olivei and Tenreyro use quarterly data from 1959 to 2005 to estimate the vector autoregression

$$y_t = B_{0,Q(t)} + B_1 \cdot t + \sum_{k=1}^{4} A_{k,Q(t)} y_{t-k} + \varepsilon_t \tag{12}$$

with y_t a 4×1 vector containing log GDP for quarter t, the log of the GDP deflator, the log of the commodity price index, and the Federal Funds rate. The calendar quarter of

[9]Olivei and Tenreyro report that the BEA is the source of the GDP and the GDP deflator series, and that the Commodity Research Bureau is the source of the commodity price index. The full dataset used by Olivei and Tenreyro is available through the AER's website.

period t is given by $Q(t)$, so each equation has 69 different unknown regression coefficients and is estimated with 144 total observations. To test that the coefficients $B_{0,j}$ and $A_{k,j}$ are equal across j for any one of the equations in (12) requires imposing 51 constraints, giving $q/n \sim 0.35$. This ratio is large enough that the naive F-test could over-reject.

The null hypothesis of no seasonal effects in equation i can be written formally as

$$\begin{aligned} H_o: \qquad & B_{0,1}^{(i)} = B_{0,m} \qquad && m = 2, \ldots, 4 \\ & A_{k,1}^{(ij)} = A_{k,m}^{(ij)} \qquad && m = 2, \ldots, 4, \quad j = 1, \ldots, 4, \quad k = 1, \ldots, 4 \end{aligned}$$

with $B_{0,m}^{(i)}$ the ith element of the vector $B_{0,m}$, and $A_{k,m}^{(ij)}$ the (i,j) element of the matrix $A_{k,m}$. To test this hypothesis, we calculate $\hat{F}$ and $\hat{G}$ and the Wald test, using the restricted VAR to estimate the excess kurtosis of the errors. These statistics are presented in Table 4.

These tests support Olivei and Tenreyro's results. Our new test statistic rejects at the 5% level for the equations with GDP Deflator, the commodity price index, and the Federal Funds Rate as the dependent variables. However, one-period-ahead GDP may not be subject to these seasonal effects—the new statistic, $\hat{G}$, has a p-value of 0.107 and so fails to reject. Notice that the naive F-test has a p-value of 0.098 and so it rejects.

Although the p-values for our new test and the naive F-test are similar, those of the Wald test are very different and much smaller. In the GDP equation, for example, its p-value is 0.035. This difference supports one of the conclusions of our Monte Carlo study—that the Wald test is more prone to over-reject than other test statistics. Note that the Wald test would lead to similar test results as the other statistics in this particular application, though, since all reject the null hypothesis at the 5% level in three of the four equations studied.

In macroeconomic applications, like Olivei and Tenreyro's, the desire to flexibly model the dynamics of the economy leads researchers to use vector autoregressions that have many unknown coefficients. It is especially difficult to accurately estimate these models with the datasets available, but it is not clear that a smaller model would be able to capture the dynamics of interest. This paper's theory suggests that one can reliably test the significance of these coefficients even if they are estimated imprecisely, and this section's analysis indicates that failing to account for the complexity of these models may give misleading results, though this does not seem to be the case in Olivei and Tenreyro's study. Finally, until we verify that the test based on $\hat{G}$ is appropriate in time series regressions with lagged dependent variables, this section's results should be viewed as preliminary.[10]

[10]Moreover, there are potential issues with stationarity that we have not dealt with. GDP might best enter the equation as its percent change and GDP deflator, the price index, and the interest rate could be modeled with multiple unit roots as well. The inclusion of a deterministic trend obviously violates our assumption on the eigenvalues of the design matrix, but could be removed if the series were otherwise stationary by

4.2 Cross-Country Growth Regressions

Our second application looks at the literature on the causes of economic growth. Over long periods of time, the benefits of a high growth rate dominate other determinants of a region's welfare, so understanding the factors that cause economic growth is important. Interest in these factors has led researchers to estimate equations of the from

$$\text{growth}_j = \beta_0 + \beta_1 x_j + \varepsilon_j \tag{13}$$

with growth_j the average rate of per capita GDP growth in country j between two specified years and x_j a vector of country-level explanatory variables. A concern in this literature is that there are many potential variables that cause economic growth, so the dimension of x_j can be large. In practice, researchers often select a small subset of those predictors and test the smaller model; this approach makes it hard to compare studies and hard to know the importance of any one variable while controlling for the effects of all of the others.

Levine and Renelt (1992) propose one solution to these problems; they use a variation of Leamer's Extreme Bounds Analysis to explain average growth rate from 1960 to 1989. Levine and Renelt estimate (13) using different subsets of the regressors and label the relationship between the ith variable and economic growth "fragile" if any two of the estimated coefficients for this variable have different signs. To make this approach computationally feasible, they restrict the subsets of the regressors they consider. Each regression includes four variables—a measure of initial per-capita income in 1960, primary school enrollment in 1960, the investment share of GDP in 1960, and the average annual population growth rate—arguing that these variables have broad support and are included in most prior empirical studies. Permutations of the other regressors are then examined, subject to the constraint that at most three additional variables enter the equation. Levine and Renelt find, perhaps unsurprisingly, that the relationship between growth and most other variables is "fragile," that there are different subsets of additional regressors for which the sign of almost any estimated coefficient switches.

Sala-i-Martin (1997) and Sala-i-Martin et al. (2004) argue that this approach is too strict. Sala-i-Martin (1997) proposes a model-averaging approach instead, one that Sala-i-Martin et al. (2004) build on, and finds that many of these variables are strongly correlated with economic growth. Sala-i-Martin also splits the regressors into a set of three that are included in every regression (replacing population growth and the investment share of GDP with the country's life expectancy in 1960) and estimates each possible regression that includes the

detrending the individual series. We ignore these issues here to make our analysis as similar as possible to the baseline model in the original.

other variables, again imposing a limit to speed up computation. Instead of comparing the two most extreme point estimates, though, Sala-i-Martin uses the entire empirical distribution of the estimated coefficients to determine the relationship between that regressor and economic growth. Sala-i-Martin et al. (2004) use a related Bayesian procedure and emphasize the posterior distributions of the regression coefficients. Both studies find that many of these regressors are correlated with growth, contradicting Levine and Renelt (1992).

In this section, we take a much different perspective: that this is a conventional estimation and testing problem. The relationship of interest is

$$\text{growth}_j = \beta_0 + \beta_w' w_j + \beta_z' z_j + \varepsilon_j \tag{14}$$

with w_j the three determinants of growth singled out by Sala-i-Martin (1997) and z_j the additional explanatory variables of interest. If $\beta_z = 0$, these additional predictors are not correlated with growth but if $\beta_z \neq 0$ they are, so it is natural to test this restriction. Although this simple analysis does not tell us any details about how the elements of z_j are related to growth, as the original studies aim to, it does support one set of conclusions over the other: if the variables z_j are jointly significant, we should not treat their relationship with growth as "fragile," and if they are insignificant there may be very little relationship to explain. We also test the hypothesis $\beta_w = 0$ to determine whether the favored variables are correlated with growth after controlling for the others.

We use Sala-i-Martin et al.'s (2004) dataset for this analysis. It includes data on economic growth from 1960 to 1996 for 88 countries, and includes 67 other country level variables, giving $k/n = 0.77$. The vector w_j includes an intercept, the enrollment rate in primary education in 1960, the level of GDP per capita in 1960, and the life expectancy in 1960, the three variables that Sala-i-Martin (1997) included in all of his regressions. Please see Sala-i-Martin et al. (2004) for a full description of the countries and variables contained in this dataset.

The test of the main hypothesis, that the additional regressors do not help explain economic growth, is rejected at the 10% level, supporting Sala-i-Martin's (1997) and Sala-i-Martin et al.'s (2004) conclusions over Levine and Renelt's 1992. It is somewhat surprising that the correction estimate, $\hat{v}$, is not further from one—it is estimated to be 0.97—given the number of restrictions tested. The F-test rejects the null hypothesis as well, agreeing with the corrected statistic. One would not normally be confident in the F-test here, but its close agreement with $\hat{G}$ should give us some confidence in this result. We also see that the p-value for the Wald test is extremely small, as we saw in the previous subsection. This is, again, consistent with our simulations that indicate the Wald test is prone to over-reject.

The statistics are presented in Table 5. The second test, for the significance of the "consensus" variables that Sala-i-Martin (1997) includes in every regression, fails to reject at 10%. While these variables could have an important structural relationship, they do not seem to have strong partial correlations with growth, and the data do not seem to justify favoring these regressors over the others. For this hypothesis, the test statistics $\hat{F}$ and $\hat{G}$ have similar values: the null hypothesis imposes only three restrictions, so the degree of correction, $\hat{v}$, is small.

5 Conclusion

Often researchers are concerned that using too large a model will bias their results—that they will find spurious and nonexistent patterns in a dataset simply because the model has many unknown parameters. This paper shows that the naive F-test has a tendency to over-reject for models with many parameters. However, this tendency can be understood and modeled, and this paper derives a new statistic that controls for model size and yields a valid test for regression models with many coefficients. Our theory suggests that this correction is especially important when the number of restrictions being tested is large, when the regressors are fat-tailed, and when the regression errors have high excess kurtosis—when those conditions are not met, both the original F-test and our corrected version are reliable. This paper's Monte Carlo evidence suggests that the F-test can over-reject in finite samples, and our empirical exercises demonstrate that the F-test and our new statistic may give different answers in practice when the original F statistic is near the test's critical values. This paper also shows that the Wald test can be unreliable when the regression model is large and should be avoided when possible.

The asymptotic theory underlying this new statistic builds on and extends similar results for the F-test in the ANOVA literature. The statistic that we present has several advantages over the ANOVA test statistics, the most important of which is its proximity to the F-test in situations where the F-test performs well. In that light, we also suggest that the statistic $\hat{G}$ also be used for homoskedastic ANOVA when the number of groups is large and the number of observations per group is small.

A Proofs and Additional Results

Lemma A.1. *Suppose the conditions of Lemma 2.1 hold. Then*

$$\frac{\boldsymbol{\varepsilon}_n' P_n^* \boldsymbol{\varepsilon}_n - q_n \sigma^2}{\operatorname{var}(\boldsymbol{\varepsilon}_n' P_n^* \boldsymbol{\varepsilon}_n \mid \mathbf{X}_n)^{1/2}} \xrightarrow{d} N(0,1) \tag{15}$$

and

$$\frac{\boldsymbol{\varepsilon}_n' (I_n - P_n) \boldsymbol{\varepsilon}_n - (n - k_n) \sigma^2}{\operatorname{var}(\boldsymbol{\varepsilon}_n' (I_n - P_n) \boldsymbol{\varepsilon}_n \mid \mathbf{X}_n)^{1/2}} \xrightarrow{d} N(0,1). \tag{16}$$

Proof of Lemma A.1. The proofs of (15) and (16) are identical, so we only present the proof of (15). Observe that

$$q_n^{-1/2}\big(\boldsymbol{\varepsilon}_n' P_n^* \boldsymbol{\varepsilon}_n - q_n \sigma^2\big) = q_n^{-1/2} \sum_{t=1}^n (\varepsilon_{n,t}^2 - \sigma^2) P_{n,tt}^* + q_n^{-1/2} \sum_{t=1}^n \sum_{s \neq t} \varepsilon_{n,t} \varepsilon_{n,s} P_{n,st}^*. \tag{17}$$

Since the errors form an independent sequence with mean zero, these two sums are uncorrelated and it suffices to prove that each term is individually asymptotically normal.

The proof that $q_n^{-1/2} \sum_t (\varepsilon_{n,t}^2 - \sigma^2) P_{n,tt}^*$ is asymptotically normal is immediate. Each summand is independent and each $P_{n,tt}^*$ is bounded between zero and one, as P_n^* is a projection matrix. Since $\varepsilon_{n,t}$ has bounded rth moments and $q_n \to \infty$, the summation satisfies the Lindeberg-Feller central limit theorem.

The proof for

$$q_n^{-1/2} \sum_{t=1}^n \sum_{s \neq t} \varepsilon_{n,s} \varepsilon_{n,t} P_{n,st}^*$$

is only slightly more difficult and follows from a central limit theorem for quadratic forms developed by Hall (1984) and de Jong (1987). Define

$$\varsigma_n^2 = \operatorname{var}\left(q_n^{-1/2} \sum_{t=1}^n \sum_{s \neq t} \varepsilon_{n,s} \varepsilon_{n,t} P_{n,st}^* \mid \mathbf{X}_n \right).$$

We can assume that ς_n^2 remains uniformly positive; if not, the second term in Equation (17) vanishes and the proof is complete.

To apply de Jong's central limit theorem (Theorem 5.2 of de Jong, 1987), we must prove the existence of a sequence of numbers M_n such that $M_n \to \infty$ and the following three conditions hold.

1. $\varsigma_n^{-2} M_n^4 \max_{s=1,\dots,n} \sum_{t \neq s} \left(q_n^{-1/2} P_{n,st}^* \right)^2 \to 0$ in probability.
2. $\max_{s=1,\dots n} \mathrm{E}\left(\varepsilon_{n,t}^2 1\left\{ |\varepsilon_{n,t}| > M_n \right\} \right) \to 0$

3. $\varsigma_n^{-2} q_n^{-1} \lambda_{\max}(P_n^* - \Lambda_n)^2 \to 0$ in probability, with Λ_n the diagonal matrix with elements $(P_{n,tt}^*)$.

Since P_n^* is idempotent, $\sum_{t \neq s} P_{n,st}^{*2} = P_{n,tt}^* - P_{n,tt}^{*2}$ almost surely, which is in turn less than one. The first condition, then, is satisfied for any $M_n = o(q_n^{1/4})$. The second condition is satisfied automatically because $\varepsilon_{n,t}$ has bounded rth moments and $M_n \to \infty$. Finally,

$$| \lambda_{\max}(P_n^* - \Lambda_n)| \leq |\lambda_{\max}(P_n^*) - \lambda_{\min}(\Lambda_n)| \leq 1$$

by construction, ensuring that the third condition is met. $\square$

Proof of Lemma 2.1. Under the null hypothesis, we have

$$\begin{aligned}\sqrt{q_n}(\hat{F}_n - 1) &= \frac{q_n^{-1/2} \boldsymbol{\varepsilon}_n' P_n^* \boldsymbol{\varepsilon}_n}{(n-k_n)^{-1} \boldsymbol{\varepsilon}_n' (I_n - P_n) \boldsymbol{\varepsilon}_n} - q_n^{1/2} \\ &= \frac{q_n^{-1/2} \left[\boldsymbol{\varepsilon}_n' P_n^* \boldsymbol{\varepsilon}_n - (c_n + o(1)) \boldsymbol{\varepsilon}_n' (I_n - P_n) \boldsymbol{\varepsilon}_n \right]}{(n-k_n)^{-1} \boldsymbol{\varepsilon}_n' (I_n - P_n) \boldsymbol{\varepsilon}_n}.\end{aligned}$$

Straightforward calculations give the mean and variance of the numerator:

$$\mathrm{E}\left(\boldsymbol{\varepsilon}_n' P_n^* \boldsymbol{\varepsilon}_n - c_n \boldsymbol{\varepsilon}_n' (I_n - P_n) \boldsymbol{\varepsilon}_n \mid \mathbf{X}_n \right) = 0 \qquad a.s.$$

and

$$q_n^{-1} \operatorname{var}\left(\boldsymbol{\varepsilon}_n' P_n^* \boldsymbol{\varepsilon}_n - c_n \boldsymbol{\varepsilon}_n' (I_n - P_n) \boldsymbol{\varepsilon}_n \mid \mathbf{X}_n \right) = \eta_n^2 \qquad a.s.$$

Lemma A.1 implies that the numerator is asymptotically normal and the denominator converges in probability to σ^2, completing the proof. $\square$

Proof of Theorem 2.2. Define a sequence of random variables $\{F_n^*\}$ such that $F_n^* \sim F(q_n, n - k_n)$. Lemma 2.1 implies that

$$\sqrt{\frac{q_n}{2\,(1 + c_n)}}\,(F_n^* - 1) \xrightarrow{d} N(0,1)$$

and

$$\frac{\sqrt{q_n}}{\eta_n}(\hat{F}_n - 1) \xrightarrow{d} N(0,1)$$

as $q_n \to \infty$. As a result,

$$\frac{\sqrt{2\,(1 + c_n)\, q_n}}{\eta_n} \left(\hat{F}_n - 1 \right) \stackrel{d}{=} \sqrt{q_n}\,(F_n^* - 1) + o_p(1)$$

and convergence in distribution follows.

Now, suppose that q_n is bounded. The numerator of the F-statistic under the null is equal to

$$\varepsilon_n' P_n^* \varepsilon_n / q_n = \hat{\alpha}_n' R_n (\mathbf{X}_n' \mathbf{X}_n)^{-1} R_n' \hat{\alpha}_n / q_n,$$

with

$$\hat{\alpha}_n = [R_n(\mathbf{X}_n \mathbf{X}_n) R_n']^{-1} R_n (\mathbf{X}_n' \mathbf{X}_n)^{-1} \mathbf{X}_n' \varepsilon_n,$$

the coefficient estimates from the regression of ε_n on $\mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}R_n$. Then we can apply, for example, White's (2000) Theorem 5.12 to prove that $\hat{\alpha}_n$ is asymptotically normal, so $q_n \hat{F}_n \overset{d}{=} \chi^2_{q_n} + o_p(1)$ and, since $(q_n \hat{F}_n)^2$ is uniformly integrable, $\mathrm{var}(q_n \hat{F}_n \mid \mathbf{X}_n) - 2q_n \to 0$ as $n \to \infty$. Since $c_n \to 0$ as well, $v_n \to 1$, completing the proof. □

Proof of Lemma 2.3. Define $\mu_4 = E(\varepsilon_{n,t}^4 \mid \mathbf{X}_n)$. We know from Lemma A.1 that $\hat{\sigma}^2 \to \sigma^2$ in probability, so it suffices to prove that $n^{-1} \sum_{t=1}^n (\hat{\varepsilon}_{n,t}^4 - \sigma^4 w_{n,1}) / w_{n,2} \to \mu_4$ in probability, with

$$w_{n,1} = n^{-1} \sum_{t=1}^n \left(6P_{n,tt} - 15P_{n,tt}^2 + 12P_{n,tt}^3 - 3\sum_{s=1}^n P_{n,st}^4 \right)$$

and

$$w_{n,2} = n^{-1} \sum_{t=1}^n \left(1 - 4P_{n,tt} + 6P_{n,tt}^2 - 4P_{n,tt}^3 + \sum_{s=1}^n P_{n,st}^4 \right).$$

Since

$$\hat{\varepsilon}_{n,t}^4 = \left(\varepsilon_{n,t} - \sum_{s=1}^n P_{n,ts} \varepsilon_{n,s} \right)^4,$$

we can calculate the expected value of $\hat{\varepsilon}^4_{n,t}$:

$$
\begin{aligned}
E(\hat{\varepsilon}^4_{n,t} \mid \mathbf{X}_n) &= \mu_4 - 4\sum_s E(\varepsilon^3_{n,t}\varepsilon_{n,s} \mid \mathbf{X}_n)P_{n,st} + 6\sum_{s,u} E(\varepsilon^2_{n,t}\varepsilon_{n,s}\varepsilon_{n,u} \mid \mathbf{X}_n)P_{n,st}P_{n,ut} \\
&\quad - 4\sum_{s,u,v} E(\varepsilon_{n,t}\varepsilon_{n,s}\varepsilon_{n,u}\varepsilon_{n,v} \mid \mathbf{X}_n)P_{n,st}P_{n,ut}P_{n,vt} \\
&\quad + \sum_{s,u,v,w} E(\varepsilon_{n,s}\varepsilon_{n,u}\varepsilon_{n,v}\varepsilon_{n,w} \mid \mathbf{X}_n)P_{n,st}P_{n,ut}P_{n,vt}P_{n,wt} \\
&= \mu_4 - 4\mu_4 P_{n,tt} + 6\mu_4 P^2_{n,tt} + 6\sigma^4 \sum_{s\neq t} P^2_{n,st} - 4\mu_4 P^3_{n,tt} \\
&\quad - 12\sigma^4 \sum_{s\neq t} P_{n,tt}P^2_{n,st} + \mu_4 \sum_s P^4_{n,st} + 3\sigma^4 \sum_s P^2_{n,st} \sum_{u\neq s} P^2_{n,ut} \\
&= \mu_4 \left(1 - 4P_{n,tt} + 6P^2_{n,tt} - 4P^3_{n,tt} + \sum_s P^4_{n,st}\right) \\
&\quad + \sigma^4 \left(6P_{n,tt} - 15P^2_{n,tt} + 12P^3_{n,tt} - 3\sum_s P^4_{n,st}\right),
\end{aligned}
$$

where all of the summations are taken from 1 to n. The last equality holds because P_n is idempotent, so

$$\sum_{u\neq s} P^2_{n,ut} = P_{n,tt} - P^2_{n,ts} \quad a.s.$$

for any u, s, and t.

The result then follows from the following limits:

$$n^{-1}\sum_{t}\varepsilon_{n,t}^{4}=\mu_4+o_p(1) \tag{18}$$

$$n^{-1}\sum_{t}P_{n,tt}\varepsilon_{n,t}^{4}=\mu_4(k/n)+o_p(1) \tag{19}$$

$$n^{-1}\sum_{t}\varepsilon_{n,t}^{4}P_{n,tt}^{2}=\frac{\mu_4}{n}\sum_{t}P_{n,tt}^{2}+o_p(1) \tag{20}$$

$$n^{-1}\sum_{t}P_{n,tt}^{3}\varepsilon_{n,t}^{4}=\frac{\mu_4}{n}\sum_{t}P_{n,tt}^{3}+o_p(1) \tag{21}$$

$$n^{-1}\sum_{s,t}P_{n,ts}^{4}\varepsilon_{n,s}^{4}=\frac{\mu_4}{n}\sum_{s,t}P_{n,ts}^{4}+o_p(1) \tag{22}$$

$$n^{-1}\sum_{t}\sum_{s\neq t}P_{n,st}^{2}\varepsilon_{n,t}^{2}\varepsilon_{n,s}^{2}=\frac{\sigma^4}{n}\sum_{t}\sum_{s\neq t}P_{n,st}^{2}+o_p(1) \tag{23}$$

$$n^{-1}\sum_{t}\sum_{s\neq t}P_{n,tt}P_{n,st}^{2}\varepsilon_{n,s}^{2}\varepsilon_{n,t}^{2}=\frac{\sigma^4}{n}\sum_{t}\sum_{s\neq t}P_{n,tt}P_{n,ts}^{2}+o_p(1) \tag{24}$$

$$n^{-1}\sum_{s,t}\sum_{u\neq s}P_{n,ts}^{2}P_{n,tu}^{2}\varepsilon_{n,s}^{2}\varepsilon_{n,s}^{2}=\frac{\sigma^4}{n}\sum_{t}\sum_{s\neq t}P_{n,tt}^{2}P_{n,ts}^{2}+o_p(1) \tag{25}$$

and

$$n^{-1} \sum_{t} \sum_{s \neq t} P_{n,st} \varepsilon_{n,t}^3 \varepsilon_{n,s} = o_p(1) \tag{26}$$

$$n^{-1} \sum_{t} \sum_{s \neq t} P_{n,st}^3 \varepsilon_{n,t} \varepsilon_{n,s}^3 = o_p(1) \tag{27}$$

$$n^{-1} \sum_{t} \sum_{s \neq t} P_{n,tt}^2 P_{n,st} \varepsilon_{n,t}^3 \varepsilon_{n,s} = o_p(1) \tag{28}$$

$$n^{-1} \sum_{s,t} \sum_{u \neq s} P_{n,ts}^3 P_{n,ut} \varepsilon_{n,s}^3 \varepsilon_{n,u} = o_p(1) \tag{29}$$

$$n^{-1} \sum_{t} \sum_{s \neq t} \sum_{u \neq s,t} P_{n,st}^2 P_{n,ut} \varepsilon_{n,t} \varepsilon_{n,s}^2 \varepsilon_{n,u} = o_p(1) \tag{30}$$

$$n^{-1} \sum_{t} \sum_{s \neq t} \sum_{u \neq s,t} P_{n,st} P_{n,ut}\, \varepsilon_{n,t}^2 \varepsilon_{n,s} \varepsilon_{n,u} = o_p(1) \tag{31}$$

$$n^{-1} \sum_{t} \sum_{s \neq t} \sum_{u \neq s,t} P_{n,tt} P_{n,st} P_{n,ut} \varepsilon_{n,t}^2 \varepsilon_{n,s} \varepsilon_{n,u} = o_p(1) \tag{32}$$

$$n^{-1} \sum_{s,t} \sum_{u \neq s} \sum_{v \neq s,u} P_{n,ts}^2 P_{n,tu} P_{n,tv} \varepsilon_{n,s}^2 \varepsilon_{n,u} \varepsilon_{n,v} = o_p(1) \tag{33}$$

$$n^{-1} \sum_{t} \sum_{s \neq t} \sum_{u \neq s,t} \sum_{v \neq s,t,u} P_{n,st} P_{n,ut} P_{n,vt} \varepsilon_{n,t} \varepsilon_{n,s} \varepsilon_{n,u} \varepsilon_{n,v} = o_p(1) \tag{34}$$

$$n^{-1} \sum_{s,t} \sum_{u \neq s} \sum_{v \neq s,u} \sum_{w \neq s,u,v} P_{n,st} P_{n,su} P_{n,sv} P_{n,sw} \varepsilon_{n,s} \varepsilon_{n,u} \varepsilon_{n,v} \varepsilon_{n,w} = o_p(1) \tag{35}$$

Equations (18)–(22) are immediate. To prove that equation (23) holds, note that (after some algebra)

$$n^{-1} \sum_{t} \sum_{s \neq t} P_{n,st}^2 \varepsilon_{n,t}^2 \varepsilon_{n,s}^2 - \frac{\sigma^4}{n} \sum_{t} \sum_{s \neq t} P_{n,st}^2 = \frac{2}{n} \sum_{t=2}^{n} (\varepsilon_{n,t}^2 - \sigma^2) \sum_{s=1}^{t-1} \varepsilon_{n,s}^2 P_{n,st}^2 + \frac{2\sigma^2}{n} \sum_{s=1}^{n-1} (\varepsilon_{n,s}^2 - \sigma^2) \sum_{t=s+1}^{n} P_{n,st}^2.$$

The first term is the summation of a uniformly integrable martingale difference sequence, so it converges to zero in probability by the law of large numbers. The second term converges to zero in probability as well because $\sum_{t=s+1}^{n} P_{n,st}^2$ is uniformly bounded. Equations (24) and (25) hold through similar arguments.

The proofs of equations (26)–(35) follow similar arguments to each other, so we will only prove (26). Note that

$$n^{-1}\sum_{t}\sum_{s\neq t}P_{n,st}\varepsilon_{n,t}^{3}\varepsilon_{n,s}=n^{-1}\sum_{t=2}^{n}\varepsilon_{n,t}^{3}\sum_{s=1}^{t-1}P_{n,st}\varepsilon_{s}+n^{-1}\sum_{s=2}^{n}\varepsilon_{n,s}\sum_{s=1}^{t-1}P_{n,st}\varepsilon_{n,t}^{3},$$

where each of the terms on the right side are averages of uniformly integrable martingale difference sequences. Consequently

$$n^{-1}\sum_{t}\sum_{s\neq t}P_{n,st}\varepsilon_{n,t}^{3}\varepsilon_{n,s}=\frac{\mu_3}{n}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\varepsilon_{n,s}P_{n,st}+o_p(1)=\frac{\mu_3}{n}\sum_{s=1}^{n-1}\varepsilon_{n,s}\sum_{t=s+1}^{n}P_{n,st}+o_p(1),$$

and this last term converges to zero in probability. □

Proof of Lemma 2.5. Under (9), the numerator of $\hat{F}_n$ becomes

$$q_n^{-1}\big[\boldsymbol{\varepsilon}_n'P_n^*\boldsymbol{\varepsilon}_n+2\delta_n'(R_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}R_n')^{-1}R_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\boldsymbol{\varepsilon}_n+\delta_n'(R_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}R_n')^{-1}\delta_n\big],$$

and so

$$\begin{aligned}\frac{\sqrt{q_n}}{\eta_n}(\hat{F}_n-1)=\frac{\boldsymbol{\varepsilon}_n'P_n^*\boldsymbol{\varepsilon}_n}{\eta_n\sigma^2\sqrt{q_n}}+2\frac{\delta_n'(R_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}R_n')^{-1}R_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\boldsymbol{\varepsilon}_n}{\eta_n\sigma^2\sqrt{q_n}}&\\+\frac{\delta_n'(R_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}R_n')^{-1}\delta_n}{\eta_n\sigma^2\sqrt{q_n}}+o_p(1).&\end{aligned}$$

Lemma 2.1 ensures that the first term converges to a standard normal. The second term has mean zero and variance (conditional on $\mathbf{X}_n$) equal to

$$2q_n^{-1}\eta_n^{-2}\delta_n'(R_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}R_n')^{-1}\delta_n.$$

This variance is in turn of order less than

$$(n/q_n)\delta_n'\delta_n\,\lambda_{\max}((R_n(n^{-1}\mathbf{X}_n'\mathbf{X}_n)^{-1}R_n')^{-1})$$

in probability and converges to zero. Similarly,

$$\frac{\delta_n'(R_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}R_n')^{-1}\delta_n}{\eta_n\sigma^2\sqrt{q_n}}\sim 1$$

in probability. □

Proof of Corollary 2.6. Suppose that q_n is bounded. Then $\hat{G}_n$ behaves like $\hat{F}_n$, and standard results apply. If $q_n \to \infty$, the corollary holds as a consequence of Lemma 2.5 $\square$

References

M. Akritas and S. Arnold. Asymptotics for analysis of variance when the number of levels is large. *Journal of the American Statistical Association*, 95:212–226, Mar. 2000.

M. G. Akritas and N. Papadatos. Heteroscedastic one-way ANOVA and lack-of-fit tests. *Journal of the American Statistical Association*, 99(466):368–382, June 2004.

S. Anatolyev. Inference in regression models with many predictors. *Journal of Econometrics*, forthcoming.

T. W. Anderson. Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(1):1–36, 1976.

T. W. Anderson, N. Kunitomo, and Y. Matsushita. On the asymptotic optimality of the LIML estimator with possibly many instruments. *Journal of Econometrics*, 157(2):191–204, 2010.

D. W. Andrews and J. H. Stock. Testing with many weak instruments. *Journal of Econometrics*, 138(1):24–46, May 2007.

A. Bathke. The ANOVA F test can still be used in some balanced designs with unequal variances and nonnormal data. *Journal of Statistical Planning and Inference*, 126:413–422, 2004.

P. A. Bekker. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, 62(3):657–681, May 1994.

D. D. Boos and C. Brownie. ANOVA and rank tests when the number of treatments is large. *Statistics & Probability Letters*, 23:183–191, May 1995.

G. A. Calvo. Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398, Sept. 1983.

J. Chao, J. A. Hausman, W. K. Newey, N. R. Swanson, and T. Woutersen. Instrumental variable estimation with heteroskedasticity and many instruments. Working Paper, Mar. 2008.

J. C. Chao and N. R. Swanson. Consistent estimation with a large number of weak instruments. *Econometrica*, 73:1673–1692, 2005.

G. Chen and R. A. Lockhart. Weak convergence of the empirical process of residuals in linear models with many parameters. *The Annals of Statistics*, 29:748–762, June 2001.

P. de Jong. A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, 75:261–277, June 1987.

Y. Fujikoshi, V. V. Ulyanov, and R. Shimizu. *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Wiley series in Probability and Statistics. Wiley, 2010.

P. Hall. Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis*, 14(1):1–16, 1984.

C. Han and P. C. B. Phillips. GMM with many moment conditions. *Econometrica*, 74(1): 147–192, 2006.

C. Hansen, J. Hausman, and W. Newey. Estimation with many instrumental variables. *Journal of Business & Economic Statistics*, 26:398–422, Sept. 2008.

X. He and Q. Shao. On parameters of increasing dimensions. *Journal of Multivariate Analysis*, 73:120–135, Apr. 2000.

P. J. Huber. Robust regression: asymptotics, conjectures and monte carlo. *The Annals of Statistics*, 1(5):799–821, 1973.

R. Koenker and J. A. F. Machado. GMM inference when the number of moment conditions is large. *Journal of Econometrics*, 93(2):327–344, Dec. 1999.

N. Kunitomo. Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations. *Journal of the American Statistical Association*, 75(371):693–700, 1980.

R. Levine and D. Renelt. A sensitivity analysis of cross-country growth regressions. *The American Economic Review*, 82(4):942–963, Sept. 1992.

E. Mammen. Empirical process of residuals for high-dimensional linear models. *The Annals of Statistics*, 24(1):307–335, 1996.

K. Morimune. Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size. *Econometrica*, 51(3):821–841, 1983.

G. Olivei and S. Tenreyro. The timing of monetary policy shocks. *The American Economic Review*, 97:636–663, 2007.

C. D. Orme and T. Yamagata. The asymptotic distribution of the F-test statistic for individual effects. *Econometrics Journal*, 9:404–422, 2006.

C. D. Orme and T. Yamagata. A simple heteroskedasticity and nonnormality robust F-test for individual effects. *Working Paper*, Feb. 2007.

S. Portnoy. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *The Annals of Statistics*, 12(4):1298–1309, 1984.

S. Portnoy. Asymptotic behavior of M-Estimators of p regression parameters when p^2/n is large; II. Normal approximation. *The Annals of Statistics*, 13(4):1403–1417, 1985.

S. Portnoy. Asymptotic behavior of the empiric distribution of M-Estimated residuals from a regression model with many parameters. *The Annals of Statistics*, 14(3):1152–1170, Sept. 1986.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL `http://www.R-project.org`.

X. X. Sala-i-Martin. I just ran two million regressions. *The American Economic Review*, 87 (2):178–183, May 1997.

X. X. Sala-i-Martin, G. Doppelhofer, and R. I. Miller. Determinants of long-term growth: a Bayesian averaging of classical estimates (BACE) approach. *The American Economic Review*, 94:813–835, Sept. 2004.

H. Scheffé. *The analysis of variance*. John Wiley & Sons, 1959.

J. H. Stock and M. Yogo. Asymptotic distributions of instrumental variable statistics with many weak instruments. *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, 2005.

J. B. Taylor. Aggregate dynamics and staggered contracts. *The Journal of Political Economy*, 88(1):1–23, 1980.

L. Wang and M. G. Akritas. Two-way heteroscedastic ANOVA when the number of levels is large. *Statistica Sinica*, 16:1387, 2006.

A. H. Welsh. On M-processes and M-estimation. *The Annals of Statistics*, 17:337–361, Mar. 1989.

H. White. *Asymptotic theory for econometricians*. Academic Press, 2000.

V. J. Yohai and R. A. Maronna. Asymptotic behavior of M-estimators for the linear model. *The Annals of Statistics*, 7:258–268, Mar. 1979.

Size for Normal Regressors, Homoskedastic Errors

			$t(5)$				$t(30)$				Exponential			
q	k/n	n	$\hat{F}$	$\hat{G}$	N	Wald	$\hat{F}$	$\hat{G}$	N	Wald	$\hat{F}$	$\hat{G}$	N	Wald
1	0.1	100	5	3	5	5	5	5	7	6	5	3	4	5
1	0.1	500	5	4	6	5	5	5	7	5	5	4	6	5
1	0.5	100	5	4	5	5	5	5	6	6	5	4	5	6
1	0.5	500	5	4	6	5	5	5	6	5	5	4	6	5
$k-1$	0.1	100	5	4	7	6	6	5	8	7	5	4	6	7
$k-1$	0.1	500	4	4	6	6	5	5	7	7	5	5	6	6
$k-1$	0.5	100	6	5	8	15	5	5	8	15	5	4	7	14
$k-1$	0.5	500	5	5	6	13	6	6	7	14	5	5	7	13
$k/2$	0.1	100	5	5	7	6	5	5	7	6	4	4	5	5
$k/2$	0.1	500	5	5	6	6	5	5	7	6	5	5	6	6
$k/2$	0.5	100	5	5	7	11	5	5	7	11	5	4	7	11
$k/2$	0.5	500	5	5	6	10	5	5	7	10	5	5	7	10

Table 1: Simulated size for a nominal 5% test, based on 5000 simulations with homoskedastic errors. The regressors are a $k \times n$ matrix drawn from the Normal distribution and include an intercept; the null hypothesis of each test imposes q restrictions. Each column contains the size for a given test statistic and error distribution.

Size for Cauchy Regressors, Homoskedastic Errors

			$t(5)$				$t(30)$				Exponential			
q	k/n	n	$\hat{F}$	$\hat{G}$	N	Wald	$\hat{F}$	$\hat{G}$	N	Wald	$\hat{F}$	$\hat{G}$	N	Wald
1	0.1	100	6	4	5	6	5	5	6	5	6	3	4	6
1	0.1	500	5	4	5	5	5	5	6	5	5	3	4	5
1	0.5	100	5	4	5	5	5	5	7	6	6	4	5	6
1	0.5	500	5	4	6	5	5	5	6	5	5	4	5	5
$k-1$	0.1	100	8	5	7	9	6	5	7	7	9	6	7	10
$k-1$	0.1	500	10	5	7	11	6	5	7	7	12	6	7	14
$k-1$	0.5	100	9	6	8	19	5	5	8	15	11	6	9	22
$k-1$	0.5	500	10	5	7	19	5	5	6	13	12	6	7	22
$k/2$	0.1	100	8	6	7	8	5	5	7	6	9	5	7	9
$k/2$	0.1	500	9	5	6	9	5	5	7	6	11	6	7	12
$k/2$	0.5	100	8	6	9	14	5	5	7	12	10	6	8	16
$k/2$	0.5	500	9	6	7	14	5	5	6	9	11	6	7	17

Table 2: Simulated size for a nominal 5% test, based on 5000 simulations with homoskedastic errors. The regressors are a $k \times n$ matrix drawn from the Cauchy distribution and include an intercept; the null hypothesis of each test imposes q restrictions. Each column contains the size for a given test statistic and error distribution.

Size for Cauchy Regressors, Heteroskedastic Errors

			$t(5)$				$t(30)$				Exponential			
q	k/n	n	$\hat{F}$	$\hat{G}$	N	Wald	$\hat{F}$	$\hat{G}$	N	Wald	$\hat{F}$	$\hat{G}$	N	Wald
1	0.1	100	5	4	5	6	5	4	6	7	5	3	4	7
1	0.1	500	5	3	4	7	5	5	7	8	5	3	4	7
1	0.5	100	5	4	5	7	5	5	7	8	5	4	5	7
1	0.5	500	5	4	5	7	6	5	7	7	5	4	5	7
$k-1$	0.1	100	8	6	8	7	6	5	8	8	10	6	7	7
$k-1$	0.1	500	10	6	7	7	7	5	7	8	13	7	8	7
$k-1$	0.5	100	10	6	9	6	6	5	8	8	12	6	9	7
$k-1$	0.5	500	11	6	7	7	6	5	7	8	14	6	8	7
$k/2$	0.1	100	8	5	7	7	6	5	7	8	9	6	7	7
$k/2$	0.1	500	9	6	7	7	6	5	7	8	11	5	6	9
$k/2$	0.5	100	8	6	8	7	6	5	7	8	11	7	10	7
$k/2$	0.5	500	9	6	7	7	6	5	7	7	12	7	8	7

Table 3: Simulated size for a nominal 5% test, based on 5000 simulations with heteroskedastic errors. The regressors are a $k \times n$ matrix drawn from the Cauchy distribution and include an intercept; the null hypothesis of each test imposes q restrictions. Each column contains the size for a given test statistic and error distribution.

Monetary Policy Empirics

	$\hat{v}$	$\hat{F}$	$\hat{G}$	$p_{\hat{F}}$	$p_{\hat{G}}$	p_{Wald}
GDP	0.95	1.39	1.37	0.097	0.106	0.035
GDP Deflator	0.97	1.56	1.55	0.038	0.041	0.006
Commodity Index	0.98	1.82	1.80	0.009	0.010	0.000
Fed. Funds	0.78	1.79	1.62	0.011	0.029	0.000

Table 4: Statistics for equation-by-equation hypothesis tests of coefficient equality for Olivei and Tenreyro's (2007) monetary policy VAR. $\hat{F}$ is the F-statistic and $\hat{G}$ is this paper's proposed corrected statistic. $p.$ is each statistic's corresponding p-value.

Cross-Country Growth Regression

	$\hat{v}$	$\hat{F}$	$\hat{G}$	$p_{\hat{F}}$	$p_{\hat{G}}$	p_{Wald}
Main Hypothesis	0.97	1.74	1.72	0.084	0.089	0.000
Comparison	1.00	1.22	1.22	0.328	0.328	0.300

Table 5: Statistics for equation-by-equation hypothesis tests of coefficient equality for cross-country growth regressions using Sala-i-Martin et al.'s (2004) dataset. $\hat{F}$ is the F-statistic and $\hat{G}$ is this paper's proposed corrected statistic. $p.$ is each statistic's corresponding p-value.

Approximate size of F-test for Nominal Size of 10%, 5%, and 1%

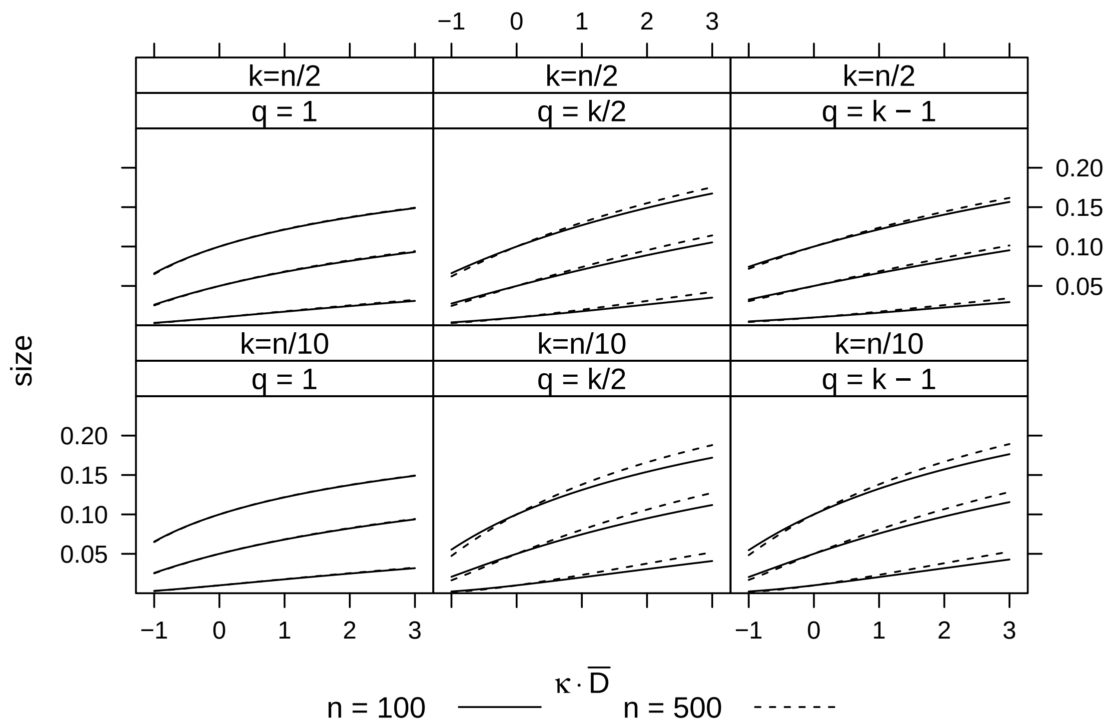

Figure 2: Approximate size of the F-test for different values of k, q, and n as a function of $\kappa\bar{D}$ (see Section 2.3 for details). Each panel presents results for $n = 100$ and $n = 500$, and for a nominal size of 1%, 5%, and 10%. The top two lines in each panel show the approximate size of the 10% test, the middle two lines the 5% test, and the bottom two the 1% test. Note that the approximate size and nominal size are equal when $\kappa\bar{D}$ is zero.